

KISHAN YADAV

Faridabad, Haryana

+91 8826495998

kishanyadav.29102002@gmail.com



Portfolio

Objective

- Full-Stack AI Engineer specializing in LLM applications, agentic workflows, and end-to-end AI system design.
- Experienced in building production-ready platforms like **ASB** using FastAPI, LangChain, custom numerical engines, multi-model LLM pipelines, and Streamlit.
- Skilled in architecting scalable AI backends, designing JSON-driven LLM subsystems, and developing automation pipelines across Machine Learning, Computer Vision, and Generative AI.
- Proficient in integrating OpenAI, Mistral/Ollama, and LLaMA-based models into structured, deterministic workflows with robust fallback strategies.
- Passionate about building reliable, user-centric, and high-impact AI products for real-world applications.

Education

Aravali College of Engineering & Management

B.Tech in Computer Science (AI & ML)

Aug 2020 – Aug 2024

Faridabad, Haryana

Shree Ram Model School

Senior Secondary (12th)

Apr 2019 – Apr 2020

Faridabad, Haryana

Experience

AI Engineer — Full-Stack AI & LLM Systems

RoboAIAPaths

Sept 2025 – Present

Faridabad ,Haryana /On Site

- Architected and developed **ASB**, a production-grade AI numerology platform featuring a modular FastAPI backend, secure API gateway, and branded Streamlit frontend.
- Engineered the **complete** numerology computation engine powering ASB: core calculations, yearly/monthly/daily predictions, profession traits, health diagnostics, and rule-based special-number logic.
- Built a provider-agnostic **LLM subsystem** (OpenAI, Ollama, Mock) with strict JSON schemas, deterministic prompt design, structured AI flows, and resilient fallback heuristics.
- Implemented SaaS-ready architecture including **feature gating**, API key security, environment-based config, caching, and structured exception handling for scalable deployment.
- Created automated PDF pipelines (ReportLab), custom matplotlib visualizations, and unified AI + numeric advisory report workflows used across the ASB platform.
- Collaborated on UI/UX, product flow, and branding to deliver a polished, reliable, AI-driven numerology advisory product with **brand-aligned Streamlit UI/UX**.

Data Science Intern → Junior Data Scientist

Global Source Technology

Sept 2024 – Aug 2025

Nashik ,Maharashtra/Remote

- Supported the analytics team during the initial part-time phase by performing EDA, preprocessing, and data cleaning on structured datasets using **Pandas** and **NumPy**.
- Assisted in developing and evaluating supervised ML models such as **Logistic Regression**, **Random Forest**, and **XGBoost** using **Scikit-learn** for internal evaluation tasks.
- Created clear visual reports and trend analyses using **Matplotlib** and **Seaborn** to support data-driven insights and internal presentations.
- Collaborated with senior data scientists to refine feature engineering workflows, applying concepts practiced in academic projects like EcoVision and MockMate.
- Gained hands-on exposure to real-world ML pipelines, documentation standards, and version control, progressing into a junior-level role toward the later phase of the internship.
- Simultaneously completed a 6-month Full Stack Data Science & AI program (Jan–Jun 2025), strengthening advanced ML, DL, and industry practices.

Technical Skills

Languages: Python

Tools & Platforms: VS Code, Jupyter Notebook, PyCharm, Spyder, Git, GitHub, MS Excel, Tableau, Streamlit Cloud, Ollama

Cloud & Big Data: Microsoft Azure ML, Azure ML Pipelines, Databricks, PySpark

Database: MySQL

Libraries & Frameworks:

- **Data Science & ML:** NumPy, Pandas, Seaborn, Matplotlib, Scikit-learn, XGBoost, LightGBM, SciPy
- **Deep Learning:** TensorFlow, PyTorch, Keras, Theano
- **Computer Vision:** OpenCV, MediaPipe, YOLO, Haar Cascade Classifier
- **NLP:** SpaCy, NLTK, Gensim, Hugging Face Transformers
- **LLMs & Generative AI:** LangChain, LangGraph, Hugging Face Transformers, GPT-3/4, Mistral-7B (Ollama), LLaMA 3, Gemini, Prompt Engineering, BERT, ChatGPT, AutoGen, Transformers Agents
- **Web & GUI:** Streamlit, Gradio, Tkinter
- **Automation & Others:** PyAutoGUI, Pycaw, SpeechRecognition, BeautifulSoup
- **Backend & Full-Stack AI:** FastAPI, REST APIs, Modular Multi-Module Architecture, API Gateway Design, SaaS Feature Gating, Caching Layers, Environment-Driven Configuration, Secure API Key Management, Structured Error Handling, Async I/O, Render Deployment
- **Visualization & Reporting:** Matplotlib (custom figure generation), ReportLab (PDF automation), Custom AI+Numeric Report Pipelines, Brand-aligned Streamlit UI/UX

MLOps: MLflow, GitHub Actions, Azure ML Pipelines, CI/CD Pipelines — Model Versioning, Experiment Tracking, Training Pipeline Automation, Model Registry, Lifecycle Management

LLMops: LangSmith, LangGraph — Prompt Evaluation, LLM Observability, Agentic Flow Management, Chain Debugging, Multi-Agent Orchestration

Concepts & Techniques:

- **Machine Learning (Supervised):** Linear & Logistic Regression, KNN, SVM, SVR, Naive Bayes, Decision Trees, Random Forest, Gradient Boosting, PCA
- **Unsupervised Learning:** K-Means, Hierarchical Clustering, DBSCAN
- **Deep Learning Architectures:** CNN, RNN, ANN, LSTM
- **NLP Techniques:** Tokenization, Lemmatization, NER, Word Embeddings, Sentiment Analysis, Text Classification
- **Transformers & Agentic AI:** Prompt Engineering, Context-Aware Chains, Retrieval-Augmented Generation (RAG), LLM Agents, Tool Use with LangChain
- **Generative Models:** GANs, VAEs
- **Data Science Workflow:** EDA, Feature Engineering, Data Preprocessing, Hyperparameter Tuning, Regularization (L1/L2)

Soft Skills: Analytical Thinking, Problem Solving, Adaptability, Collaboration, Communication, Research-Oriented

Projects

ASB – Full-Stack AI Numerology Platform | *FastAPI, Streamlit, LangChain, OpenAI/Ollama, ReportLab*

- Built a production-grade AI numerology platform with a modular FastAPI backend, LLM-powered interpretation engine, and branded Streamlit frontend.
- Engineered the complete numerology computation engine including mystical triangle, yearly/monthly/daily predictions, profession traits, health diagnostics, and rule-based special-number system.
- Developed a provider-agnostic LLM subsystem (OpenAI, Ollama, Mock) using strict JSON schemas, deterministic prompts, and robust fallback heuristics.
- Created automated PDF generation workflows (ReportLab), custom matplotlib visualizations, and unified AI+numeric reporting pipelines.
- **Impact:** End-to-end SaaS-ready platform combining deterministic numerology logic with AI-driven insights.

MockMate: Virtual Mock Interview Assistant | *Python, Streamlit, LangChain, LangGraph, Ollama* GitHub

- Built an AI-powered mock interview platform generating contextual questions using Mistral-7B via LangChain.
- Integrated voice input, transcription, and real-time scoring through prompt-based feedback pipelines.
- Implemented structured feedback, result export (PDF/JSON), and session-wise progress analysis.
- **Client Approach:** Designed to help job seekers practice interviews with AI-generated coaching and evaluation.

StyloGenie – AI Wardrobe Assistant | *LangChain, Streamlit, Mistral, Pexels API* GitHub

- Developed a LangChain agent recommending outfits based on item, color, and occasion using memory-aware prompts.
- Applied color theory for contextual styling and fetched visuals dynamically using the Pexels API.
- Deployed in a responsive Streamlit interface with real-time outfit rendering.
- **Client Approach:** Personalized styling aid for fashion platforms, stylists, and e-commerce wardrobe curation.

EcoVision – Intelligent Trash Sorting | *TensorFlow, Deep Learning, OpenCV* GitHub

- Trained MobileNetV2-based classifier on TrashNet dataset to detect 5+ waste categories with 90%+ accuracy.
- Applied image augmentation and optimization techniques to improve generalization and robustness.
- Built a real-time video classification pipeline using OpenCV for smart bin applications.
- **Client Approach:** Positioned as a solution for smart city projects, IoT-driven waste management, and sustainability tech.

DreamSpaceAI – Room Redesign with GenAI | *SDXL, Hugging Face, T2I Adapter, Python* GitHub

- Implemented a depth-aware image-to-image pipeline using Stable Diffusion XL with T2I Adapter for style transfer.
- Used Hugging Face diffusers for GPU-accelerated rendering of room redesigns based on user-defined themes.
- Enabled multimodal input (image + voice/text) and exportable outputs with real-time preview.
- **Client Approach:** Interior design previews, architectural mockups, and creative visualization for client engagement.

Certifications

NPTEL – Soft Skills

Jul 2023 – Oct 2023



Naresh i Technologies – Full Stack Data Science and AI

Jan 2025 – Jun 2025

